www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

# Applied Machine Learning

## Data-Centric Computing

March 2017

# Applied Machine Learning - DCC

- Data Mining and Knowledge Discovery for Data-Centric Systems

**Aloja**
Benchmarking Predictive Analytics

- Modeling and Prediction
- Anomaly Detection
- Recommenders

**Hi-EST**
Cognitive Self-management

- Reinforcement Learning
- Data Streams

**PanCancer**
Mining Genomic Patterns

- Knowledge Discovery
- Association Rules

**AIS**
Behavior Modeling

- Time-Series Characterization
- Cognitive Computing

**PASI**
Analytics as a Service

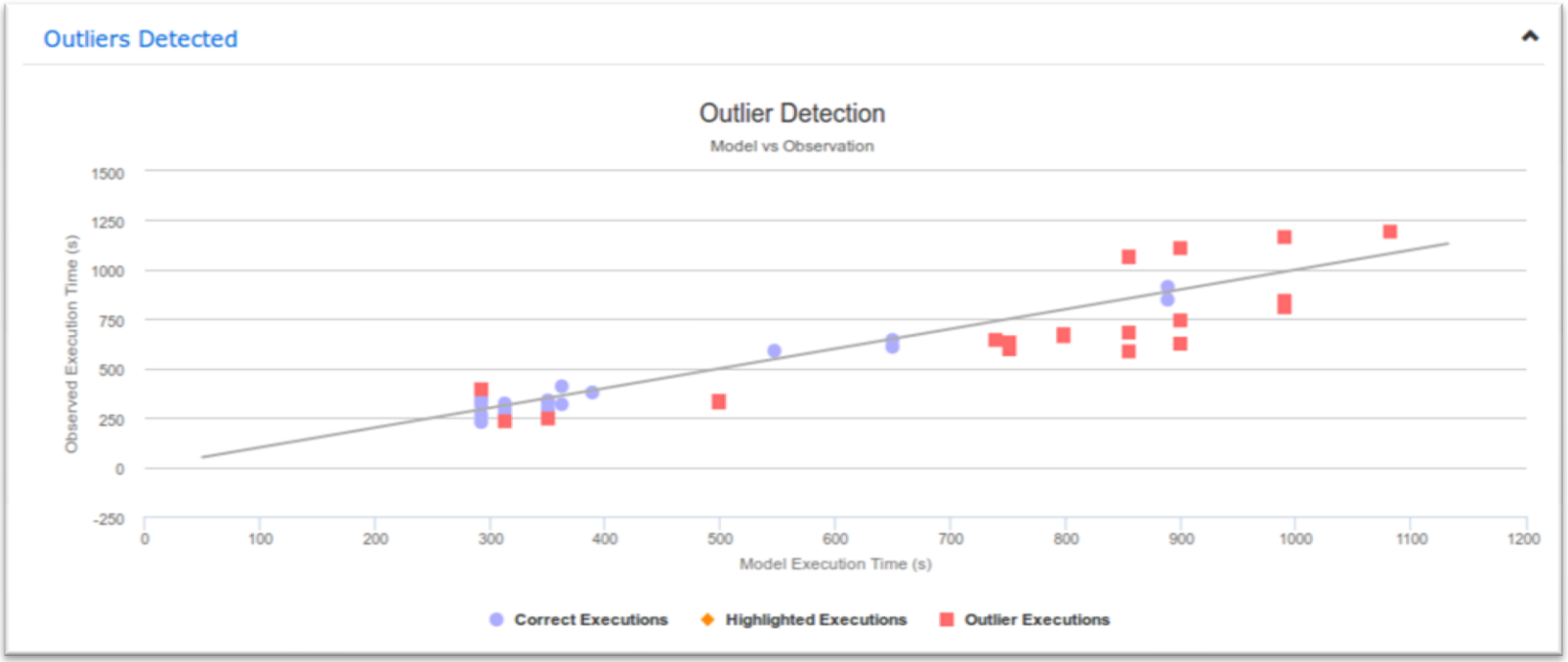
- Tools for Predictive Analytics
- Supercomputing Environments

**LightHouse**
Edge Analytics

- Edge / Fog Computing
- In-DataBase Analytics
- Anomaly Detection

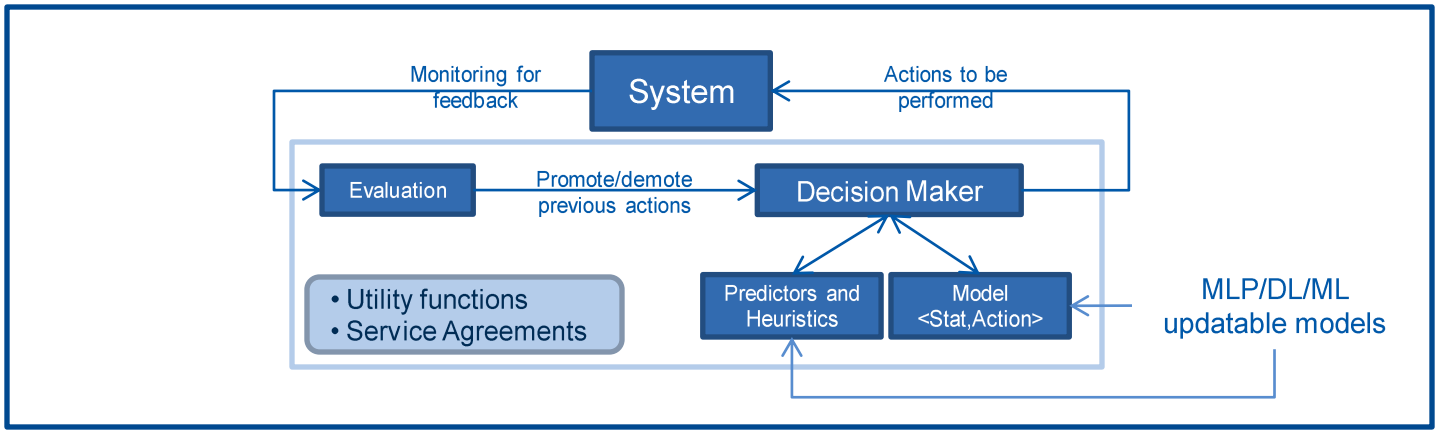
# ALOJA-ML: Benchmarking Analytics

- Analyze Big-Data jobs, deployments and HW
  - Predict behavior elements → Extract knowledge
    - Modeling and Prediction for "Big Data" jobs behavior
    - Model-based + Data-driven Anomaly Detection
    - Recommendation-guided Benchmarking
    - Knowledge Discovery from Information

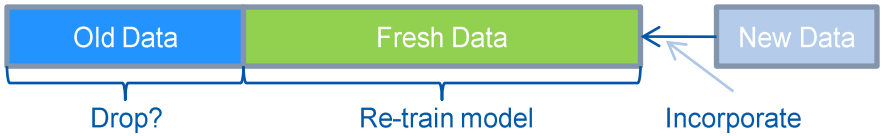


# Hi-EST – Learning, Planning and Deciding

- Scheduling and Planning in heterogeneous scenarios
  - Unconventional HW for modeling:
  - Adaptive/reinforcement learning methods:
    - Q-learning, Neuro-Dynamics and State-Action learning

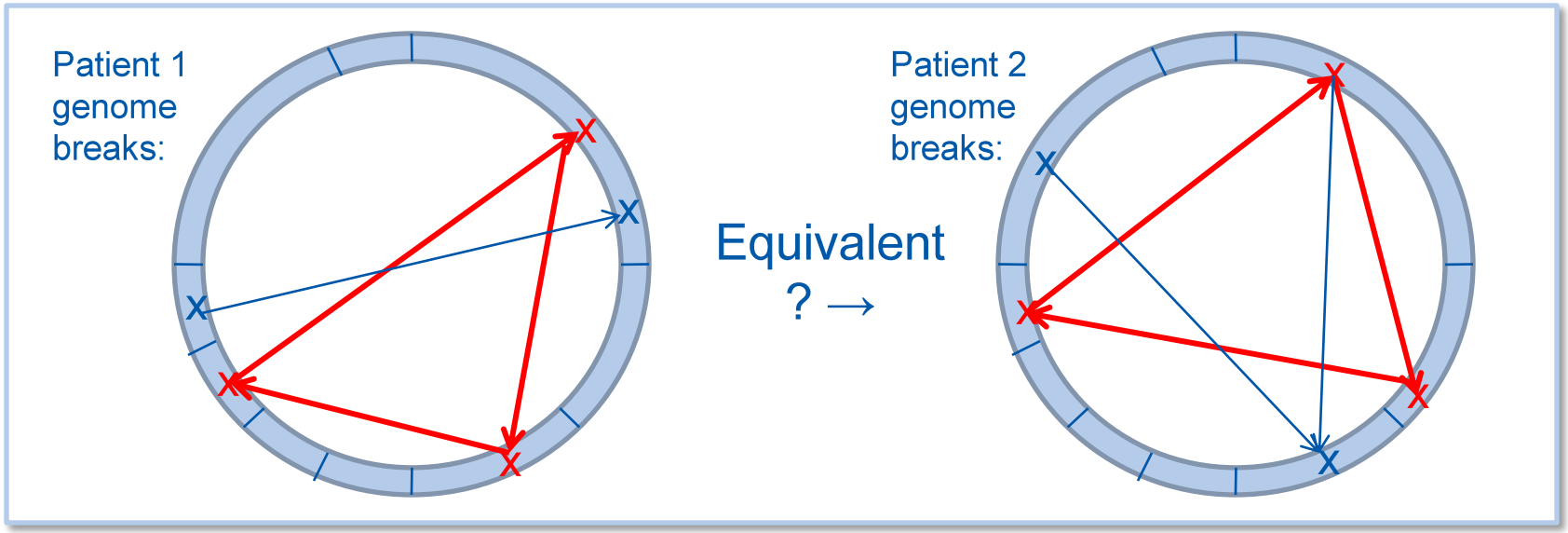



  - Data-Streaming:
    - On-line learning and Time Series
    - CRBMs and Hidden Markov Models



# PanCancer – Mining chromosome breaks

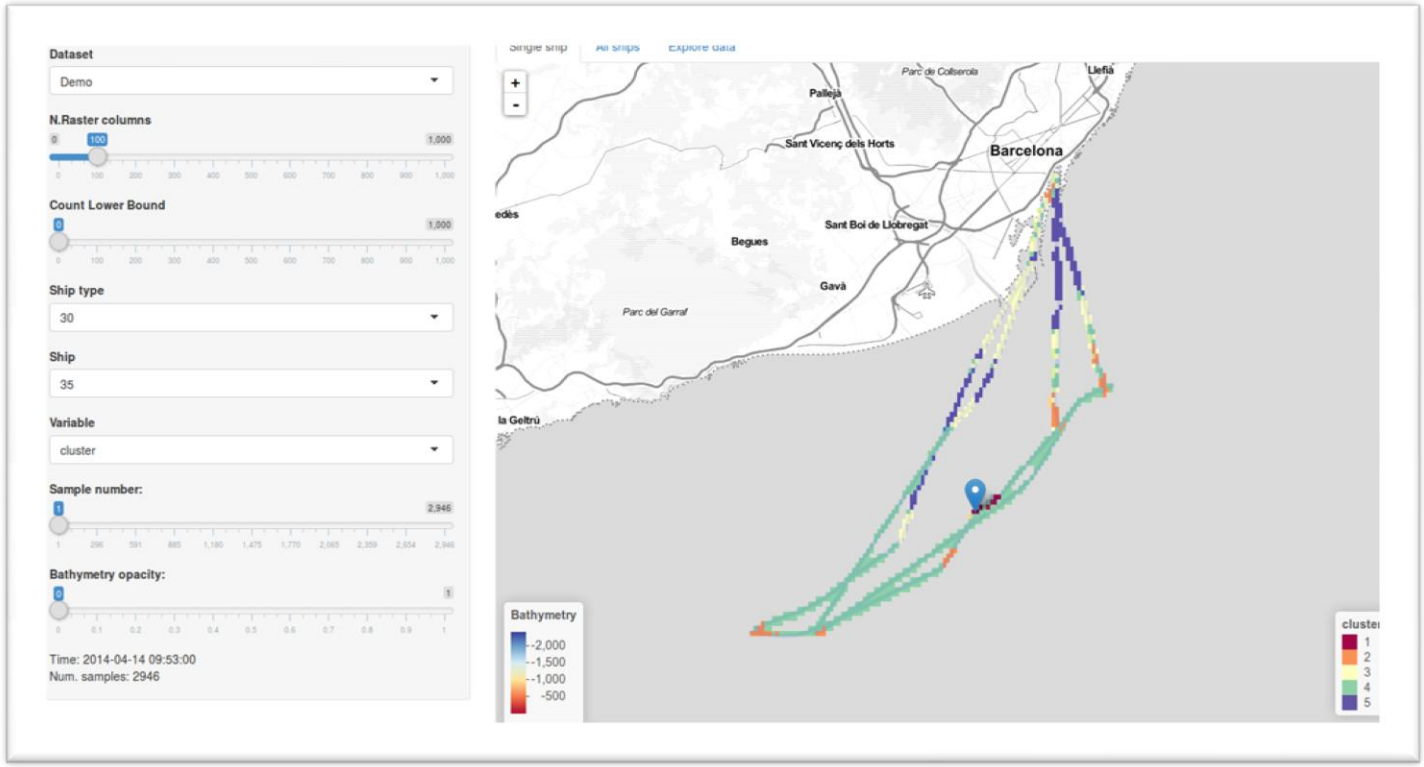
- Genome Patterns
  - Compilation of (Big) Data and treatment for easy exploration
  - Pattern mining and knowledge extraction




- Applied technologies
  - Methods using pattern mining, like Association Rules
  - Also using distributed architectures for heavy processing (Spark, etc.)

# AIS – Ships Data Mining

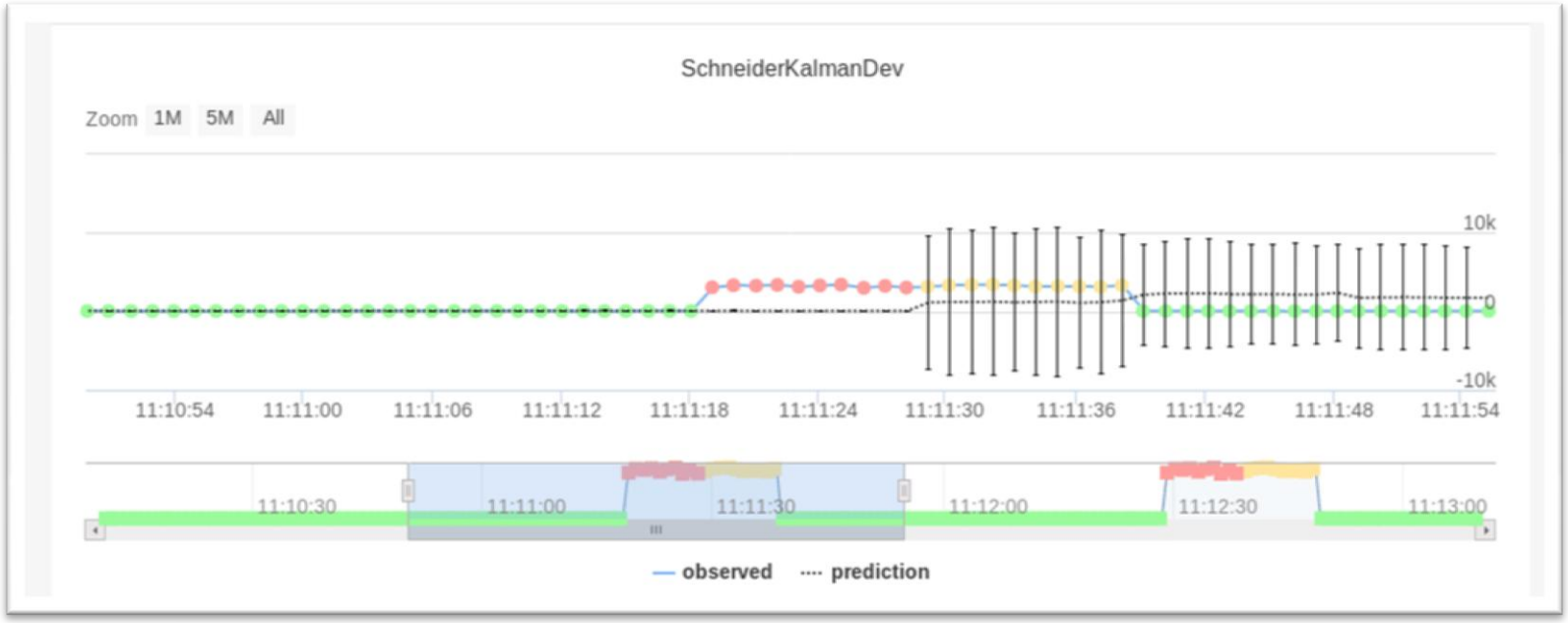
- Ship Behavior Patterns
  - Automate characterization of Ships from AIS traces
  - Relate patterns with environmental features (emissions,pollution, …)




- Applied technologies
  - Methodologies using “Cognitive Computing” (Deep Learning, CRBMs, DBNs…)

# LightHouse – Analytics on the Edge/Fog

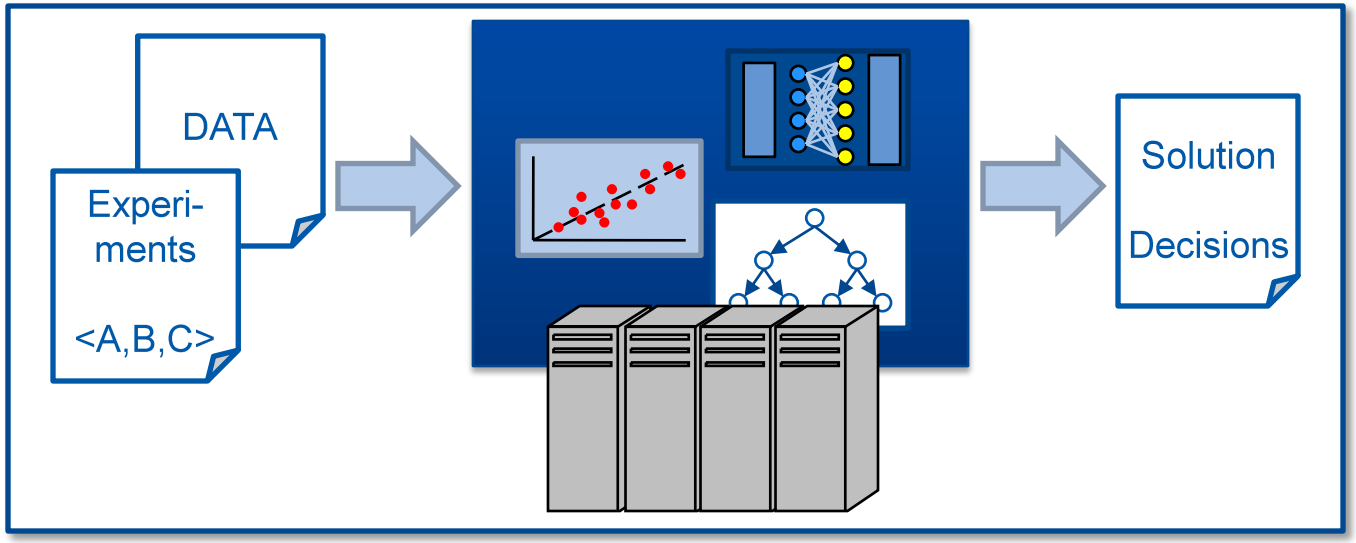
- Edge and Fog Computing
  - Fog-Computing sensor and application networks
    - Internet of Things
    - Anomaly detection and Forecasting
      - E.g. Kalman Filters (Time-Series)
  - In-DataBase computing



# PASI – Pred. Analytics in Supercomputers

- Predictive Analytics in Supercomputing Infrastructures
  - Modeling and Prediction
  - Anomaly Detection
  - Characterization
  - Recommendation

- Automation of Data Pipelines
  - Without requiring expert deployment
  - Offering direct methods instead of individual functionalities



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

# THANKS FOR YOUR ATTENTION